MirrorBot: Using Human-inspired Mirroring Behavior To Pass A Turing Test

Mihai Polceanu

▶ **To cite this version:**

Mihai Polceanu. MirrorBot: Using Human-inspired Mirroring Behavior To Pass A Turing Test. IEEE Conference on Computational Intelligence and Games, Aug 2013, Niagara Falls, Canada. pp.201 —-208. hal-00864153

HAL Id: hal-00864153

https://hal.archives-ouvertes.fr/hal-00864153

Submitted on 23 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MirrorBot: Using Human-inspired Mirroring Behavior To Pass A Turing Test

Mihai Polceanu
Université Européenne de Bretagne, Ecole Nationale d'Ingénieurs de Brest,
Lab-STICC, Centre Européen de Réalite Virtuelle
Technopôle Brest-Iroise, 29238 Brest Cedex 3, France
E-mail: polceanu@enib.fr

***Abstract*—Believability of automated characters in virtual worlds has posed a challenge for many years. In this paper, the author discusses a novel approach of using human-inspired mirroring behavior in MirrorBot, an Unreal Tournament 2004 game bot which crossed the humanness barrier and won the 2K BotPrize 2012 competition with the score of 52.2%, a record in the five year history of this contest. A comparison with past contest entries is presented and the relevance of the mirroring behavior as a humanness improvement factor is argued. The modules that compose MirrorBot's architecture are presented along with a discussion of the advantages of this approach and proposed solutions for its drawbacks. The contribution continues with a discussion of the bot's results in humanness and judging accuracy.**

I. Introduction

The goal of a Turing test [1], as described by its creator in 1950, is to determine whether a computer program is able to act like a human being. The test does not impose restrictions on the inner-workings of the computer program, and leaves room for implementations which are not conscious but exhibit superficial intelligence. The original Turing test requires a virtual chat room where communication is done via text messages, in which a participant in the discussion is judged as being a human or a robot. This participant passes the Turing test if the interlocutor cannot deny that the conversation was carried out with a human.

A. Evaluation environment

The 2K BotPrize[1] competition [2] is a game bot variant of the Turing test, which replaces chatting with a shooting game environment. The competition was started in 2008 and has gathered teams of participants from various countries to develop bots that can fool judges into thinking they are actually human players. Entries gradually improved each edition, finally achieving behavior indistinguishable from that of humans in 2012 through two bots: UT^2 [3] and MirrorBot.

As the BotPrize competition is based on Unreal Tournament 2004 (UT2004), a first person shooter (FPS) game, the way players interface with the environment induces in-game behavioral differences between humans and bots. These differences originate in the fact that humans control an avatar via keyboard and mouse for movement and use a visual display to understand the environment. Contrarily, bots advance using movement vectors and have access to a navigation graph to find paths. Visual feedback from the game enables humans to easily bypass obstacles, wait for and properly board elevators that stand in their path. Bots however, have trouble with obstacles and are usually associated with saccaded movement, which rarely appears in the case of human players. Because movement and aim are handled separately by humans (keyboard and mouse respectively), the navigation and shooting directions of human players do not usually coincide. Furthermore, because human players generally have slower reaction times than automated characters, expert players who have high accuracy and significantly better overall performance can be mistaken for bots, by casual players. For example, in some games, top players are accused of cheating because to their excellent skills.

Undoubtedly, differences between human and bot players extend beyond the way they perceive and act in the environment, but interface-related issues constitute the starting ground for developing human-like bots.

B. The role of the judge

The first two editions of BotPrize were based on series of gaming sessions in which a human, a bot and a judge played together. The role of the human and the bot was to play against each other, including attacking the judge, while the judge's role was to decide if the other two players are human or robot.

To make the environment more suited for evaluating human-like behavior, since 2010 a more flexible version of this feature was introduced [4], which made the judges part of the game as opposed to being only observers. The 2010 edition of BotPrize brought a modification (mod) to UT2004 that was based on a special judging gun (adapted functionality of the original link gun) with which human judges were able to express their votes regarding other players. A judging decision consisted in tagging a player as being a human or a bot, the result of which was the death of the target if the judgment was correct, and the death of the judge otherwise.

An update of this mod was introduced in the 2012 edition, removing the instant penalty of wrong decisions and allowing judges (and also bots) to change their vote during a game. The newly proposed design was based on observed drawbacks of former editions and stemmed from the concept of making the judge role part of the game.

Due to the aggressive nature of the original game, when two players engage in combat they usually tend to avoid directly facing their enemy, which leads to short confrontation times. The newly added judging role triggers the appearance

[1] 2K BotPrize contest website http://www.botprize.org/

978-1-4673-5311-3/13/$31.00 ©2013 IEEE

of new behavior in the game. The focus on other players' behavior is increased and interaction becomes friendlier as judges tend to spend extra time to study each opponent in order to spot actions that are generally attributed to bots.

Moreover, with chatting disabled, communication attempts can be performed through movements and shooting patterns. A judge may prompt a fellow player to communicate and expect a behavioral response that seems intelligent or resembles some form of meaning. The way in which a player responds to the judge's queries may significantly incline the decision towards a favorable (human-like) or inconvenient (bot-like) result.

C. A focus on realtime interaction

To seem human-like, it is important for a bot to have a global coherence of actions. However, being able to demonstrate human traits when interacting with the judges greatly improves its believability. Taking into account the constraints imposed by Unreal Tournament 2004 and the BotPrize judging mod, the approach in this work focuses on realtime interaction between MirrorBot and the players it encounters during the game.

The relatively short interaction time, which is due to the rapid nature of the game, puts a greater importance on the awareness of an encountered player. Bots which are inattentive in such situations may loose credibility in the eyes of a judge. In contrast, when judges receive meaningful responses from the bot, their decision can be positively influenced.

The subject of this paper is the implementation of MirrorBot, which passed the humanness barrier in the 2012 edition of the 2K BotPrize contest with a score of 52.2% humanness (defined as how human-like a player seems to the judges). The novel design feature of MirrorBot consists in the use of mirroring behavior that enables it to react, in a meaningful manner, to complex context-dependent situations in realtime.

II. Related work

During the five editions of the BotPrize competition participants have had a range of approaches, some of which achieved record levels of human-like behavior which guaranteed their winning. An overview of the scores achieved by participants throughout the history of the contest is illustrated in Fig. 1. Some of the entries have participated in more than one contest edition, exposing their evolution in the humanness score.

To account for the different evaluation methodology used in the first two editions where a rating of 4 out of 5 convinced judges was required to win the major prize as opposed to 50% humanness in later editions, a score scaling was applied to merge all results into the same chart. Thus, these scores were scaled to equalize the major prize winning threshold, using the scaling rule: $humanness = \frac{rating*50}{4}\%$. For example, the AMIS[2] bot won the minor prize in the 2008 edition with a mean rating of 2.4 out of 5 convinced judges which corresponds to roughly 30% humanness within the new contest rules.

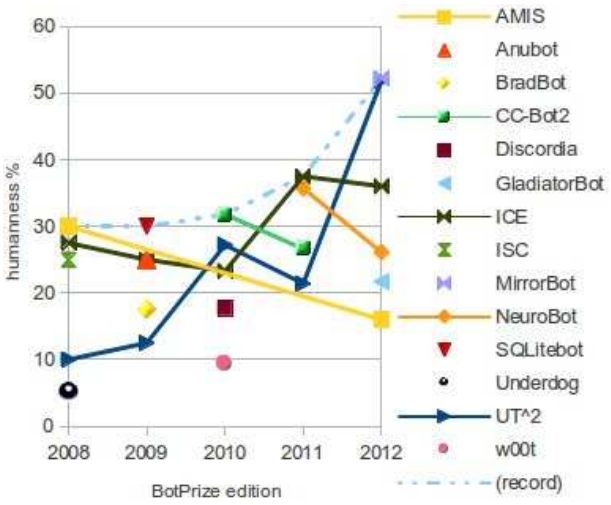

Fig. 1. Humanness scores for bot entries throughout the contest history

A. Design approaches in BotPrize

The implementation approaches of BotPrize participants can be classified into two categories represented by classical rule-based systems and those based on neural networks. Both categories harbor record holders, without exhibiting a clear pattern of which category is more fitted to achieve human-level believability. However, the top performance of the solutions submitted in each contest edition increased gradually.

The rule-based category contains the majority of entries, including AMIS (winner in 2008), SQLiteBot[3] (winner in 2009), CC-Bot2 [5], [6] (winner in 2010) and MirrorBot (winner in 2012). In the context of FPS games where reaction time is essential, rule-based architectures contain robust modules which can be optimized for high performance, giving them a design-wise speed advantage. Moreover, this type of approach can be systematically tested to remove bugs or unwanted behavior. However, to achieve high level functionality, especially when attempting to behave in a human-like fashion, the architecture tends to grow in complexity and may require increased development time and therefore most attempts sacrifice complex behavior to avoid the development effort of implementing many particular cases.

The second category of contest entries relies on training neural networks used to model the way in which humans play. This category is represented by NeuroBot [7], ICE [8] (winner in 2011) and UTˆ2 [3], [9] (winner in 2012). The advantage brought by this type of approach consists in the capability of these systems to exhibit complex behavior which they previously learned from extensive human recordings. However, the drawback is that a significant amount of time is spent on acquiring the required data. Furthermore, to achieve such complex traits the neural network can grow to such extent that it becomes a challenge for it to run in realtime. Authors of NeuroBot reported issues when they were obliged to run their bot on a machine with lower performance. In the case of UTˆ2 the network grown through neuroevolution had to be constrained with rules in order to remove undesired behavior.

[2]Michal Stolba, AMIS bot https://artemis.ms.mff.cuni.cz/pogamut/tiki-index.php?page=Botprize%202008%20winning%20bot

[3]Jeremy Cothran, SQLiteBot http://aigamedev.com/open/article/sqlite-bot/

In order to benefit from the decision speed advantage, MirrorBot was built upon a rule-based architecture which allows computationally cheap behavior generation and requires no prior learning stages. The bot's architecture is based on a small number of well defined modules which can be activated depending on context. One of these modules is entirely devoted to the requirement to behave in a believable manner and consists in a human-inspired mirroring behavior which allows it to exhibit complex behavior based on realtime observation instead of a preprocessing stage.

B. Achieving believability

Believability has been assessed for virtual characters in several contexts such as conversational agents [10], [11] where facial expressions and gestures are important, while in the case of emotional agents, the focus is set on reacting to the state of mind of others [12]. Due to the fact that the meaning of believability varies depending on the context in which it is evaluated, assessment of this aspect in the case of Unreal Tournament should rely on a set of behavioral traits. Such criteria has been described in [13] where a set of ten traits are proposed to measure believability of a human-controlled character in a virtual environment.

Another description of believability, from the perspective of how bots are detected as non-human, is given in [4] where the author lists, based on the comments of human judges, a set of features or mechanisms, which made the judges decide that the bots were not in fact human. These features include very accurate shooting, forgetting about opponents and lack of awareness. Based on the comments of the judges and implementation details described in papers about the mentioned entry bots, a list of required mechanisms for believability in the BotPrize context has been compiled in Table I. These mechanisms each contribute in a certain amount to the bots' level of believability.

TABLE I. MECHANISMS TO ACHIEVE BELIEVABLE BEHAVIOR

Label	Mechanism
A	casual stops during gameplay
B	dodging skill when engaged
C	evasion when in disadvantage
D	hiding ability to trick enemy
E	hold grudge on enemy (focus on one enemy)
F	imitation (offline, from recordings)
G	imitation (in real-time)
H	judging ability
I	long-term memory (places/events)
J	limited firing accuracy
K	resource gathering based on needs
L	situation awareness (observe players' actions)
M	smooth trajectory of movement
N	target chasing
O	weapon choice depending on context

Although this list of mechanisms is dependent on the context proposed by the BotPrize, a link can be drawn to existing literature (Table II). The requirements to which the mechanisms are linked are described in [13], and have been chosen as a suitable link due to their applicability to FPS games. Some features, although implemented atomically, may exhibit more than one believability trait. For example, the hiding ability (D) can be implemented as a procedure in the bot, but can considered as contributing to the reactiveness (B1) of the bot and its planning (B7) competency. Likewise, a believability trait can have several implementations, for example perception (B6) is achieved through situation awareness (L) and smooth trajectories (M) as performed by humans using their visual perception system.

TABLE II. MAPPING CONTEXT-DEPENDENT FEATURES ONTO MORE GENERIC BELIEVABILITY TRAITS (ADAPTED FROM [13])

Requirements for believable virtual characters	Bot features (mechanisms) in the BotPrize context
[B1: Reaction]	B C D L
[B2: Reaction time]	J
[B3: Variability]	A B
[B4: Unpredictability]	A B
[B5: Understandable]	D E G H
[B6: Perception]	L M
[B7: Planning]	D K
[B8: Memory]	E I

To evaluate in what proportion do the winning bots use such mechanisms to achieve believable behavior, they are compared based on the list in Table I. Information in Table III was compiled from available source code[4,5,6] and previously cited papers [3], [6], [14], [15] which discuss the implementation of the bots. Some features may exist in the bots' implementation without having been published and therefore may not appear in this table. In the case of the 2008 and 2009 editions, judging was not available (n/a) for bots.

TABLE III. FEATURES OF WINNING BOTS

	AMIS (2008)	SQLiteBot (2009)	CC-Bot2 (2010)	ICE (2011)	UT^2 (2012)	MirrorBot (2012)
A	✓				✓	✓
B	✓	✓	✓	✓	✓	✓
C		✓	✓	✓		✓
D		✓		✓		
E	✓		✓		✓	✓
F				✓	✓	
G						✓
H	n/a	n/a		✓	✓	✓
I		✓			✓	
J	✓	✓	✓	✓	✓	✓
K	✓	✓	✓	✓	✓	✓
L			✓	✓	✓	✓
M			✓	✓	✓	✓
N	✓	✓	✓	✓	✓	✓
O	✓	✓	✓	✓	✓	✓

The comparison shows an increase of features included in the winning entries as the contest progressed. Some features are shared by all contest entries such as dodging skills, accuracy limitation, resource gathering, target chasing and weapon choice, representing the most basic subset of competence for a bot. Other characteristics appear scattered and appear to have a small relative improvement to the perceived level of humanness. However, three of these features only appear in the design of the last three winners - the judging ability, offline and online imitation abilities. The appearance of the judging ability was triggered by the new design of the game since 2010, but was not sufficient to obtain a level of humanness over 50%. Therefore, the introduction of imitation triggered a

[4]Michal Stolba, AMIS source code
http://artemis.ms.mff.cuni.cz/pogamut_files/amisbot-botprize2008.jar

[5]Jeremy Cothran, SQLiteBot source code
http://code.google.com/p/sqlitebot/

[6]Jacob Schrum, Igor V. Karpov, UT^2 source code
http://www.cs.utexas.edu/users/nn/downloads/software/UT2-Botprize2012.zip

score increase from 31.8% in 2010 to 52.2% in 2012. While the winner of the 2011 edition and the co-winner of 2012 used offline imitation, based on learning on human records, MirrorBot differentiates itself with the use of realtime mimicry.

III. Relevance of mirroring

Latest studies show that human beings, as well as species of primates and birds, are able to perform mirroring, in their interaction with other individuals. The discovery, in macaque monkeys, of neurons that fire when an individual observes actions performed by others in the environment [16] has posed a challenge for further research to determine the role of these neurons in the animals' behavior. The name "mirror neuron" was coined for neurons which have the property of firing both when an individual performs an action and when it observes another perform the same action or a similar one. In a thorough review of existent work and results in this field [17], the authors make reference to several links between mirror neurons and adjacent fields of research, highlighting the importance of imitation in clinical purposes, intention understanding and social cognition.

Imitation and mirror processes may constitute a foundation for social behavior such as interacting with and relating to other members of a community. In [18] it is suggested that mirroring might provide a basis for social connectedness and mutual acknowledgement between individuals, but the authors mention that, in humans, basic imitation is gradually replaced by more constructive interaction such as reciprocation, joint attention to objects and identification of self with others. Another trait associated with mirroring is mutual recognition [18] which expresses shared understanding between individuals. In a social context, nonconscious behavioral mimicry, called the "chameleon effect" [19], was shown to increase affiliation between people and through evolution, it has gained a role in social interaction [20]. Although imitation is covered in a wide range of research, two main senses are identified in [21], i.e. the capacity to replicate an observed behavior and to learn new behavior through observation. While future research of mirror neurons may elucidate whether mimicking is an innate ability for humans or it is acquired through experience [22], this type of behavior plays an important role in interhuman relations.

In the context of the BotPrize competition, where the goal of the competing bots is to act like humans as much as possible, due to the fact that humans are social beings it is only natural for them to expect other human players to engage in social interaction with them. As a result of the introduction of judging for both humans and bots in the contest, aggressiveness has been partially replaced by a focus on friendly interaction between players, and thereby social interaction is fostered.

Most studies in the field of neuroscience focus on the role of mirroring from the perspective of the individual who performs it, but in the competition context it is more meaningful to explore what the mirrored human perceives from the bot's behavior. It has been observed, from studies in the fields of social psychology and child development, that the initiators of actions generally react in a positive way to being imitated [23]. Herein, authors present several imitation experiments which involved both children and adults, showing an increased likelihood for the individual who is imitated to create a social bond with the imitator. Being imitated was found to be more rewarding in the case of children, while adults were less comfortable with other adults that performed high-fidelity imitation which was interpreted as mockery or plagiarism.

In a FPS game, mirroring can be performed on character movement, gaze or shooting direction, fire pattern, choice of weapon and sequences of jumps. Mirroring such behavior has not extensively been performed in these contexts, but there exist a number of implementations using virtual characters. Such an approach which used virtual humanoids that exhibited socially relevant facial expressions is presented in [24], with results showing that participants were more receptive to virtual characters that showed communicative intentions. Gaze congruency was approached in [25] where authors used an interactive eye-tracking paradigm in which participants could interact with a virtual character that followed the gaze direction of its user in real-time. Their results show that the perceived level of humanness increases with higher fidelity of gaze reactions. The process of imitation recognition was investigated in [26], where the authors observed that the bilateral extrastriate body area (EBA) plays a major role in the ability of humans to recognize action congruency when they imitate or are imitated.

Due to the commonness of mirroring, humans expect this kind of behavior from others. Moreover, they tend to assume a certain level of confirmation and understanding when they are mirrored. The main design principle of MirrorBot is to exploit this assumption human players make, in order to avoid unmasking the bot's lack of consciousness.

Given evidence of the existence of a mirroring mechanism in humans and its effects, the implementation of MirrorBot focused not on simulating the internal workings of this process, but on exhibiting this behavior when the bot came in contact with other players. Based on both psychology research and on gameplay common sense, the bot was designed on the assumption that if the mirrored player is a human, MirrorBot would inherit the level of humanness when it replays the behavior back to its originator. Not only the replayed behavior is intrinsically as human as its model, but also the process of mirroring has origins in common human behavior.

IV. Implementation of MirrorBot

The design of MirrorBot is based on two modules which are interchanged during gameplay. The architecture is divided into two main modules, default and mirroring, due to the fact that the former generates behavior while the latter acts only as a playback activity. An overview of these main behaviors and their components is illustrated in Fig. 2.

The first module implements default behavior such as graph-based and ray-based navigation, target choice, aiming, shooting, opponent awareness and judging. This behavior offers a basis for playing the game and all contest entries contain versions of these functionalities, as they represent a basic interface with the game environment. The way in which these functionalities are implemented however, may significantly influence outcome of the judgment because glitches in navigation for instance can quickly destroy the illusion of humanness [27],

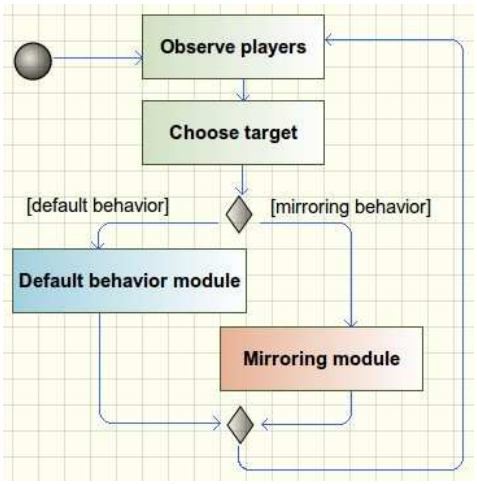

Fig. 2. Overview of MirrorBot's behavior

[28]. An additional module that gave the bot its name is the mirroring module which, when activated, records the actions of one target and then, after a short time delay, replays them from the bot's perspective. This secondary module is activated only in the case when MirrorBot encounters a friendly player. Each module is described in the following subsections.

A. Default behavior

Inspired from how human players control their avatar in the game, i.e. via keyboard and mouse, the implementation of MirrorBot's default behavior fosters independence between aim control and navigation (Fig. 3). Likewise, shooting is triggered when the bot's orientation provides a high chance of damaging an enemy, while aiming is responsible for picking locations in the environment that require attention.

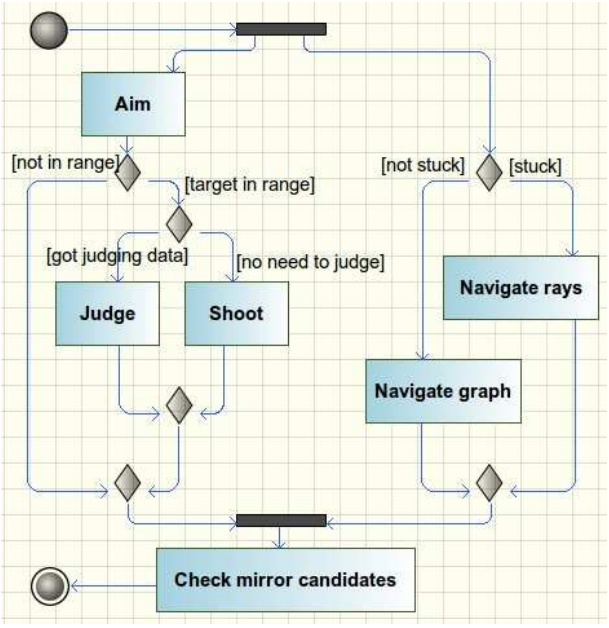

Fig. 3. MirrorBot's default behavior

1) Observe players: The central input of the bot provides information of other players' actions. Shooting direction of other players is especially studied in order to isolate predator-prey pairs and aggressive or friendly shooting patterns. MirrorBot's decision making process takes into account a player's aggressiveness; i.e. if that enemy is shooting and the bullets are directed towards itself. This module influences target choice, mirror module activation and evasive behavior.

2) Aim: The implementation of the aiming module is based on successively adjusting the bot's orientation to a given focus location. The focus is changed to the current target, which is decided upon considering the level of its aggressiveness and a time-limited grudge. Once assigned, a target will be kept until the grudge expires. The grudge is replenished if the current enemy replies with fire power. Furthermore, in the case of a moving target, the focus is not on the enemy's current location, but rather set to a future location computed using the enemy's speed. If no target is acquired, the focus is aimed to a point computed through a linear interpolation of the next two navigation targets, determined by the navigation module. This feature is inspired from the ability of humans to anticipate their trajectory, especially at sharp corners in the map. This feature enables the bot to emerge from a corner while looking in the correct direction, in case enemies appear.

3) Shoot: The shooting module is responsible for choosing between primary and secondary fire modes and which weapon to use, based on the efficiency of the weapon and its available ammunition. A special case of shooting is handled when the bot has changed its decision on the humanness of a player. In this case, it will use the special judging gun to tag that player. To increase the fidelity of the judging function, the bot uses the caused damage sensor to decide if its target has been hit.

4) Dodging: With the purpose of self defense, based on inputs of the observation module, the bot avoids the sight of its enemies. If in combat, the bot will calculate the shooting direction of its enemy, and will move to the closest position that allows it to evade direct fire. This holds true for all players that are currently shooting, so that the bot avoids moving in the fire direction of another player.

5) Navigate: In the unfortunate but highly possible situation where MirrorBot cannot escape a certain location due to a malfunction of the main graph navigation algorithm, it triggers the use of a secondary module based on ray casting. A total of 24 rays are used for this algorithm, divided into 16 rays that stem horizontally from the center, and 8 additional rays that stem downwards at a 45 degree angle (Fig. 4). The horizontal rays are used to detect obstacles from which the bot will move away, and the downward rays are used to detect holes or cliffs to avoid falling.

Fig. 4. Backup navigation based on ray casting

The results received from the rays represent a boolean value which signifies whether the ray has hit an object in the environment. In Figure. 4, affirmative results are drawn in red and the lack of collision is displayed in green. Ray collision in the horizontal case will cause the bot to move away from obstacles, while in the downward case avoidance is triggered by the lack of collision.

The main navigation module is an improved version of the default Pogamut [29] Loque navigator, tweaked for smoother trajectories, greedy resource gathering and additional obstacle avoidance using the same set of rays as in the backup navigation. Although improved, the algorithm still exhibits malfunctions which include missing elevator rides when other players use the lift in the same time and subtle bot-like movement. However, most of this behavior is masked in interaction situations due to the influence of dodging which introduces a small amount of unpredictability in its motion.

6) Judge: The judging mechanism of the bot is based on the opinion of the majority. Data collected by the observation module is used to record the judgments for a player by others. While it gathers more data, the bot will change its decision on each player, and when its decision changes, MirrorBot will vote the player again to adjust the tag assigned to that player.

B. Mirroring behavior

When a candidate is found suitable for performing the mirroring behavior, all default behavior is ceased and MirrorBot starts recording the full motion and shooting data of the selected player. The functions contained by the mirroring behavior are illustrated in Fig. 5.

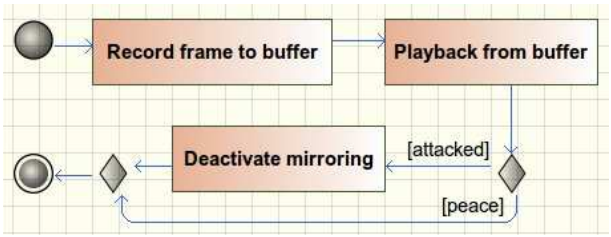

Fig. 5. MirrorBot's mirroring behavior

Candidates are selected in function of their friendliness and orientation, for instance if the player is not shooting with a dangerous weapon and if it is facing the bot. An example of choosing a candidate is shown in Fig. 6 where one of the judges commences friendly interaction with MirrorBot.

Fig. 6. Candidate for mirroring (left = MirrorBot, right = Judge)

Once the mirroring behavior is activated, player information is recorded as keyframes that contain the exact state of the targeted player at a given moment. A player state consists of the location of the target, velocity, rotation, jumping and crouching state, shooting mode and carried weapon. Keyframes are then buffered and played back with a delay of approximately 100-200 milliseconds (Fig. 7). The real delay time can be up to 500 milliseconds due to the latency introduced by the client-server connection. This value was tuned empirically to simulate a plausible reaction time needed for a humans to interpret and react to visually perceived actions [30].

Fig. 7. Performing the mirroring behavior in the 2K BotPrize 2012 contest (top = external view, bottom = judge view)

When a mirroring keyframe is executed, MirrorBot's state is changed to match the information stored in that particular keyframe. The target's location is used to establish an axis of symmetry with MirrorBot. Using this axis, all actions except forward/backward movement are recreated as seen in a mirror; i.e. target shoots to its top-right and MirrorBot shoots to its top-left. The forward and backward movements are opposite to the mirror effect so to maintain distance in the case where the target retreats or advances.

To account for the need of reactivity while mirroring, some control conditions were implemented to stop this behavior when the bot is being attacked, the mirroring target loses interest or a reasonable amount of time has passed. The maximum amount of time allowed to perform mirroring behavior for one player was empirically set to approximately 8 seconds, due to the fast nature of the game. Likewise, if the mirrored player turns away, MirrorBot will do the same, therefore if the bot does not see its target for a time limit of approximately 3 seconds, the module will deactivate. Even though compared with the total length of the game the mirroring module is activated briefly, the fact that it is used especially during encounters with players gives it a higher visibility.

A possible scenario which would cast MirrorBot in disadvantage is mirroring another bot, especially if that bot is stuck or is performing very simplistic behavior. Stuck or inactive bots may seem friendly and therefore would fit all conditions for

a good mirroring candidate. A third player may consider that MirrorBot is actually a bot because it has the same behavior of the stuck bot. To ameliorate this disadvantage, the behavior exhibited by the mirroring module checks if the targeted player is too silent, and if several frames are detected in which there is no significant movement, slight random movements are introduced to produce some diversity in the playback. However, the problem is not completely solved, as the target may exhibit movement but in a bot-like manner. Coping with this problem should include knowledge on whether the target is a bot or a human so to increase the probability that mirroring is activated in case the target is human.

The original goal of the game is to win by the number of players that one has virtually killed (frags). This focus on interaction may significantly decrease the frag number, causing a decrease in game-oriented performance of MirrorBot, however in the context of BotPrize frags were not taken into account in calculating the humanness level.

Playing back actions in realtime allows the bot to exhibit highly complex behavior in correct contexts. Continuity in the sequence of actions is enforced by the time delay which enables a reasonable number of frames to be buffered before execution, so that interruptions are less likely to occur. In an attempt to use communication to test whether a player is human or not, judges may draw patterns using weapon fire, perform systematic left-right movements or any other combination of movements that they find appropriate and understandable by others. If a player responds accordingly to this kind of communication attempts, a human may assume that the player's behavior was the result of a conscious interlocutor.

V. RESULTS

Evaluation of the contest entries consisted in calculating the number of times the judges voted a player as being human, divided by the total number of times that player has been voted. Although bots were given the ability to judge, only the votes of humans were taken into account for the results. Based on information found on the BotPrize[7], each player was judged approximately 25 times, thus conferring high reliability to the results. Tables IV and V contain the humanness scores of humans and bots respectively, in the 2012 edition of BotPrize. The average scores are calculated overall and the number of votes made by each judge may vary.

TABLE IV. HUMANNESS SCORES FOR HUMAN JUDGES (BOTPRIZE 2012)

#	Player	Humanness
1	Samaneh Rastegari	53.3 %
2	Craig Speelman	52.2 %
3	John Weise	30.8 %
4	Chris Holme	26.3 %
	average	*41.4 %*

Humanness was defined as a measurement of how human-like a player seems to a judge. Fast movement and high accuracy, associated with bots, can be exhibited also by expert players, who can be mistakenly judged as being robots.

Results indicate that mimicry has played an important role in the judges' perception of humanness. While offline mimicry enables a bot to achieve high credibility in movement patterns, roughly the same high score was achieved by MirrorBot using only realtime mirroring. This highlights the believability improvement brought by this technique in the UT2004 environment, but experiments with mirroring in environments beyond the FPS games may elucidate whether this behavioral trait can be extrapolated to other contexts. One may postulate that given the fact that the principle of mirroring consists in replaying certain observed actions, it could be applied in a larger range of informed virtual environments.

TABLE V. HUMANNESS SCORES FOR BOT ENTRIES (BOTPRIZE 2012)

#	Player	Humanness
1	MirrorBot	52.2 %
2	UT^2	51.9 %
3	ICE-CIG2012	36.0 %
4	NeuroBot	26.1 %
5	GladiatorBot	21.7 %
6	AmisBot	16.0 %
	average	*34.2 %*

Although the accuracy of judging did not represent a requirement for winning the contest, for the purpose of evaluating the judging behavior of MirrorBot, the results of the correctness of each player's decisions are presented in Table VI.

TABLE VI. JUDGING ACCURACY OF EACH PLAYER (BOTPRIZE 2012)

#	Player	Accuracy
1	Chris Holme	60.9 %
2	John Weise	60.8 %
3	MirrorBot	60.0 %
4	NeuroBot	56.4 %
5	UT^2	54.6 %
6	GladiatorBot	50.5 %
7	Craig Speelman	50.0 %
8	Samaneh Rastegari	47.8 %
9	ICE-CIG2012	45.8 %
10	AmisBot	44.4 %

MirrorBot's decision mechanism based on majority belief, although simple, performed reasonably well compared to the average accuracy of other players which was approximately 52%. By its nature, this strategy is only as efficient as the correctness of other players' judgments. It also depends on whether MirrorBot encountered a specific player in the game in order to record that player's decision. This result may also be influenced by the necessity that the humans perform judgments, possibly more frequently than the bots, given by the context of the competition. Additionally, these results could have been improved by using a heuristic that would rule out possibly untrustworthy decisions made by other players, or by a more complete probabilistic model.

Moreover, an accurate knowledge of which players are, or seem, human or bot-like could ameliorate the previously described issue of mirroring the wrong players. Information gathered with the purpose of judging players can be used to increase the probability of performing the mirroring behavior with humans or even bots that are considered human by other players.

[7] 2K BotPrize contest website http://www.botprize.org/

VI. Conclusions and future work

The approach introduced in this paper was to enhance a computer game bot with a common human trait of being able to mimic peers in a social context. Reviewing related studies in the origin and importance of the mirroring phenomenon and studying what past BotPrize entries have used to achieve their humanness level, has led to a more flexible architecture with separate modules which can be tweaked individually. The improvements brought by the addition of the mirroring behavior in MirrorBot are proved by its results in the 2012 edition of BotPrize.

The current architecture, while being computationally inexpensive, allows more high-level behavior to emerge due to the activation of subsets of its components. Examples of such behavior include grudge-like behavior, bullet dodging, combat assistance based on hunter-prey relations, corner aim anticipation, enemy tauting when fleeing and following are combined, strategic position in combat based on the greedy resource gathering and, most importantly for a human-like appearance, realtime social reply ability. Because mirroring is not effective when targets do not exhibit human-like behavior, future work will focus on target selection using data collected from participants. Furthermore, an extended evaluation of mirroring as a believability enhancing factor will be addressed in different contexts and compared to offline learning.

Although the bot was programmed to work with the UT2004 environment, its mirroring capability could be successfully implemented in a wider range of informed virtual environments or games in which social interaction is endorsed, where it could be used to exhibit complex behavior with little computation time spent to achieve it.

References

[1] A. M. Turing, "Computing Machinery and Intelligence," *Mind*, vol. LIX, pp. 433–460, 1950.

[2] P. Hingston, "A turing test for computer game bots," *Computational Intelligence and AI in Games, IEEE Transactions on*, vol. 1, no. 3, pp. 169–186, sept. 2009.

[3] J. Schrum, I. Karpov, and R. Miikkulainen, "Utˆ2: Human-like behavior via neuroevolution of combat behavior and replay of human traces." in *CIG*, S.-B. Cho, S. M. Lucas, and P. Hingston, Eds. IEEE, 2011, pp. 329–336.

[4] P. Hingston, "A new design for a turing test for bots," in *Computational Intelligence and Games (CIG), 2010 IEEE Symposium on*, aug. 2010, pp. 345 –350.

[5] R. Arrabales, A. Ledezma, and A. Sanchis, "Towards conscious-like behavior in computer game characters," in *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on*, sept. 2009, pp. 217 –224.

[6] R. Arrabales, J. Muoz, A. Ledezma, G. Gutierrez, and A. Sanchis, "A machine consciousness approach to the design of human-like bots," in *Believable Bots*, P. Hingston, Ed. Springer Berlin Heidelberg, 2012, pp. 171–191.

[7] D. Gamez, Z. Fountas, and A. Fidjeland, "A neurally-controlled computer game avatar with human-like behaviour," *Computational Intelligence and AI in Games, IEEE Transactions on*, vol. PP, no. 99, p. 1, 2012.

[8] D. Hirono and R. Thawonmas, "Implementation of a human-like bot in a first person shooter: second place bot at botprize 2008," in *Proc. Asia Simulation Conference*, 2009, p. 5.

[9] I. Karpov, J. Schrum, and R. Miikkulainen, "Believable bot navigation via playback of human traces," in *Believable Bots*, P. Hingston, Ed. Springer Berlin Heidelberg, 2012, pp. 151–170.

[10] Z. Ruttkay, C. Dormann, and H. Noot, "Evaluating ecas - what and how," in *Proc. of the AAMAS02 Workshop on Embodied Conversational Agents*, 2002.

[11] E. Bevacqua, K. Prepin, R. Niewiadomski, E. de Sevin, and C. Pelachaud, "GRETA : Towards an Interactive Conversational Virtual Companion," in *Artificial Companions in Society: perspectives on the Present and Future*, 2010, pp. 1–17.

[12] W. S. Reilly, J. Carbonell, R. Simmons, W. Scott, W. Scott, N. Reilly, and N. Reilly, "Believable social and emotional agents," 1996.

[13] F. Tencé, "Probabilistic behaviour model and imitation learning algorithm for believable characters in video games," Ph.D. dissertation, Université de Bretagne Occidentale, Brest, France, 2011.

[14] Z. Fountas, D. Gamez, and A. Fidjeland, "A neuronal global workspace for human-like control of a computer game character." in *CIG*, S.-B. Cho, S. M. Lucas, and P. Hingston, Eds. IEEE, 2011, pp. 350–357.

[15] R. Thawonmas, S. Murakami, and T. Sato, "Believable judge bot that learns to select tactics and judge opponents," in *Computational Intelligence and Games (CIG), 2011 IEEE Conference on*, 31 2011-sept. 3 2011, pp. 345 –349.

[16] G. Pellegrino, L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti, "Understanding motor events: a neurophysiological study," *Experimental Brain Research*, vol. 91, pp. 176–180, 1992.

[17] G. Rizzolatti and M. Fabbri-Destro, "Mirror neurons: from discovery to autism," *Experimental Brain Research*, vol. 200, pp. 223–237, 2010.

[18] P. Rochat and C. Passos-Ferreira, "From imitation to reciprocation and mutual recognition," in *Mirror Neuron Systems*, ser. Contemporary Neuroscience, J. A. Pineda, Ed. Humana Press, 2009, pp. 191–212.

[19] T. L. Chartrand and J. A. Bargh, "The chameleon effect: the perception-behavior link and social interaction." *Journal of Personality and Social Psychology*, vol. 76, no. 6, pp. 893–910, 1999.

[20] J. Lakin, V. Jefferis, C. Cheng, and T. Chartrand, "The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry," *Journal of Nonverbal Behavior*, vol. 27, pp. 145–162, 2003.

[21] G. Rizzolatti, "The mirror neuron system and imitation," *Kodaly Envoy*, vol. 66, no. 5, pp. 557–560, 2005.

[22] G. Rizzolatti and M. Fabbri-Destro, "The mirror system and its role in social cognition," *Current Opinion in Neurobiology*, vol. 18, no. 2, pp. 179 – 184, 2008, cognitive neuroscience.

[23] S. Rosen, L. Musser, and J. Brown, "Reactions of third-graders to recognition allocation after being peer-imitated," *Current Psychology*, vol. 21, pp. 319–332, 2002.

[24] L. Schilbach, A. M. Wohlschlaeger, N. C. Kraemer, A. Newen, N. J. Shah, G. R. Fink, and K. Vogeley, "Being with virtual others: Neural correlates of social interaction," *Neuropsychologia*, vol. 44, no. 5, pp. 718 – 730, 2006.

[25] U. Pfeiffer, B. Timmermans, G. Bente, K. Vogeley, and L. Schilbach, "A non-verbal turing test: Differentiating mind from machine in gaze-based social interaction," *PLoS one*, vol. 6, p. e27591. e27591, 2011, record converted from VDB: 12.11.2012.

[26] Y. Okamoto, R. Kitada, A. Sasaki, T. Morita, S. Itakura, T. Kochiyama, H. C. Tanabe, and N. Sadato, "Like-me: The role of extrastriate body area for recognition of being imitated," *Neuroscience Research*, vol. 65, Supplement 1, no. 0, pp. S195 –, 2009, abstracts of the 32nd Annual Meeting of the Japan Neuroscience Society (Neuroscience 2009).

[27] J. Weizenbaum, "Eliza - a computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, Jan. 1966.

[28] F. Tencé, C. Buche, P. D. Loor, and O. Marc, "The challenge of believability in video games: Definitions, agents models and imitation learning," *GAMEON-ASIA'2010, France*, 2010.

[29] J. Gemrot, R. Kadlec, M. Bda, O. Burkert, R. Pbil, J. Havlek, L. Zemk, J. imlovi, R. Vansa, M. tolba, T. Plch, and C. Brom, "Pogamut 3 can assist developers in building ai (not only) for their videogame agents," in *Agents for Games and Simulations*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009, vol. 5920, pp. 1–15.

[30] R. M. Hecht D. and H. G, "Multi-modal stimulation, response time, and presence," in *Proceedings of 8th Annual International Workshop on Presence (Presence 2005), London, UK*, 2005.